



NF-kappa B sequence on step 63 was obtained from HIV-1 (nt 349-374, nt 9434-9458)

K03455. Human immunodeficiency...[gi:1906382]

Related Sequences,

Protein, PubMed, Taxonomy

LOCUS HIVHXB2CG 9719 bp ss-RNA VRL 19-AUG-1999  
DEFINITION Human immunodeficiency virus type 1 (HXB2), complete genome;  
HIV1/HTLV-III/LAV reference genome.  
ACCESSION K03455 M38432  
VERSION K03455.1 GI:1906382  
KEYWORDS TAR protein; acquired immune deficiency syndrome; complete genome;  
env protein; gag protein; long terminal repeat (LTR); pol protein;  
polyprotein; proviral gene; reverse transcriptase; transactivator.  
SOURCE Human immunodeficiency virus type 1.  
ORGANISM Human immunodeficiency virus type 1  
Viruses; Retroid viruses; Retroviridae; Lentivirus; Primate  
lentivirus group.  
REFERENCE 1 (bases 493 to 674; 9577 to 9718)  
AUTHORS Ratner,L., Haseltine,W., Patarca,R., Livak,K.J., Starcich,B.,  
Josephs,S.F., Doran,E.R., Rafalski,J.A., Whitehorn,E.A.,  
Baumeister,K., Ivanoff,L., Petteway,S.R. Jr., Pearson,M.L.,  
Lautenberger,J.A., Papas,T.S., Ghrayeb,J., Chang,N.T., Gallo,R.C.  
and Wong-Staal,F.  
TITLE Complete nucleotide sequence of the AIDS virus, HTLV-III  
JOURNAL Nature 313 (6000), 277-284 (1985)  
MEDLINE 85111123  
PUBMED 2578615  
REFERENCE 2 (bases 1 to 653)  
AUTHORS Starcich,B., Ratner,L., Josephs,S.F., Okamoto,T., Gallo,R.C. and  
Wong-Staal,F.  
TITLE Characterization of long terminal repeat sequences of HTLV-III  
JOURNAL Science 227 (4686), 538-540 (1985)  
MEDLINE 85090465  
REFERENCE 3 (sites)  
AUTHORS Allan,J.S., Coligan,J.E., Barin,F., McLane,M.F., Sodroski,J.G.,  
Rosen,C.A., Haseltine,W.A., Lee,T.H. and Essex,M.  
TITLE Major glycoprotein antigens that induce antibodies in AIDS patients  
are encoded by HTLV-III  
JOURNAL Science 228 (4703), 1091-1094 (1985)  
MEDLINE 85192537  
REFERENCE 4 (sites)  
AUTHORS Sodroski,J., Patarca,R., Rosen,C., Wong-Staal,F. and Haseltine,W.  
TITLE Location of the trans-activating region on the genome of human  
T-cell lymphotropic virus type III  
JOURNAL Science 229 (4708), 74-77 (1985)  
MEDLINE 85244627  
REFERENCE 5 (sites)  
AUTHORS Arya,S.K., Guo,C., Josephs,S.F. and Wong-Staal,F.  
TITLE Trans-activator gene of human T-lymphotropic virus type III  
(HTLV-III)  
JOURNAL Science 229 (4708), 69-73 (1985)

MEDLINE 85244626  
 REFERENCE 6 (sites)  
 AUTHORS van Beveren,C.P., Coffin,J. and Hughes,S.  
 TITLE Appendix B: HTLV-3/LAV genome  
 JOURNAL (in) Weiss,R.L., Teich,N., Varmus,H. and Coffin,J. (Eds.);  
 RNA TUMOR VIRUSES, SECOND EDITION, 2, Vol. 2: 1102-1123;  
 Cold Spring Harbor Laboratory, Cold Spring Harbor (1985)  
 REFERENCE 7 (sites)  
 AUTHORS Rosen,C.A., Sodroski,J.G. and Haseltine,W.A.  
 TITLE The location of cis-acting regulatory sequences in the human T cell  
 lymphotropic virus type III (HTLV-III/LAV) long terminal repeat  
 JOURNAL Cell 41 (3), 813-823 (1985)  
 MEDLINE 85228232  
 REFERENCE 8 (sites)  
 AUTHORS Rabson,A.B., Daugherty,D.F., Venkatesan,S., Boulukos,K.E.,  
 Benn,S.I., Folks,T.M., Feorino,P. and Martin,M.A.  
 TITLE Transcription of novel open reading frames of AIDS retrovirus  
 during infection of lymphocytes  
 JOURNAL Science 229 (4720), 1388-1390 (1985)  
 MEDLINE 85300515  
 REFERENCE 9 (sites)  
 AUTHORS Allan,J.S., Coligan,J.E., Lee,T.H., McLane,M.F., Kanki,P.J.,  
 Groopman,J.E. and Essex,M.  
 TITLE A new HTLV-III/LAV encoded antigen detected by antibodies from AIDS  
 patients  
 JOURNAL Science 230 (4727), 810-813 (1985)  
 MEDLINE 86044509  
 REFERENCE 10 (sites)  
 AUTHORS Rosen,C.A., Sodroski,J.G., Goh,W.C., Dayton,A.I., Lippke,J. and  
 Haseltine,W.A.  
 TITLE Post-transcriptional regulation accounts for the trans-activation  
 of the human T-lymphotropic virus type III  
 JOURNAL Nature 319 (6054), 555-559 (1986)  
 MEDLINE 86118720  
 REFERENCE 11 (sites)  
 AUTHORS di Marzo Veronese,F., Copeland,T.D., DeVico,A.L., Rahman,R.,  
 Oroszlan,S., Gallo,R.C. and Sarngadharan,M.G.  
 TITLE Characterization of highly immunogenic p66/p51 as the reverse  
 transcriptase of HTLV-III/LAV  
 JOURNAL Science 231 (4743), 1289-1291 (1986)  
 MEDLINE 86122937  
 REFERENCE 12 (sites)  
 AUTHORS Kan,N.C., Franchini,G., Wong-Staal,F., DuBois,G.C., Robey,W.G.,  
 Lautenberger,J.A. and Papas,T.S.  
 TITLE Identification of HTLV-III/LAV sor gene product and detection of  
 antibodies in human sera  
 JOURNAL Science 231 (4745), 1553-1555 (1986)  
 MEDLINE 86151663  
 REFERENCE 13 (sites)  
 AUTHORS Kramer,R.A., Schaber,M.D., Skalka,A.M., Ganguly,K., Wong-Staal,F.  
 and Reddy,E.P.  
 TITLE HTLV-III gag protein is processed in yeast cells by the virus  
 pol-protease  
 JOURNAL Science 231 (4745), 1580-1584 (1986)  
 MEDLINE 86151671  
 REFERENCE 14 (sites)  
 AUTHORS Lee,T.H., Coligan,J.E., Allan,J.S., McLane,M.F., Groopman,J.E. and

Essex, M.

TITLE A new HTLV-III/LAV protein encoded by a gene found in cytopathic retroviruses

JOURNAL Science 231 (4745), 1546-1549 (1986)

MEDLINE 86151661

REFERENCE 15 (sites)

AUTHORS Dayton, A.I., Sodroski, J.G., Rosen, C.A., Goh, W.C. and Haseltine, W.A.

TITLE The trans-activator gene of the human T cell lymphotropic virus type III is required for replication

JOURNAL Cell 44 (6), 941-947 (1986)

MEDLINE 86161683

REFERENCE 16 (sites)

AUTHORS Sodroski, J., Goh, W.C., Rosen, C., Tartar, A., Portetelle, D., Burny, A. and Haseltine, W.

TITLE Replicative and cytopathic potential of HTLV-III/LAV with sor gene deletions

JOURNAL Science 231 (4745), 1549-1553 (1986)

MEDLINE 86151662

REFERENCE 17 (sites)

AUTHORS Arya, S.K. and Gallo, R.C.

TITLE Three novel genes of human T-lymphotropic virus type III: immune reactivity of their products with sera from acquired immune deficiency syndrome patients

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83 (7), 2209-2213 (1986)

MEDLINE 86177573

REFERENCE 18 (sites)

AUTHORS Jones, K.A., Kadonaga, J.T., Luciw, P.A. and Tjian, R.

TITLE Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1

JOURNAL Science 232 (4751), 755-759 (1986)

MEDLINE 86179897

REFERENCE 19 (sites)

AUTHORS Sodroski, J., Goh, W.C., Rosen, C., Dayton, A., Terwilliger, E. and Haseltine, W.

TITLE A second post-transcriptional trans-activator gene required for HTLV-III replication

JOURNAL Nature 321 (6068), 412-417 (1986)

MEDLINE 86230863

REFERENCE 20 (sites)

AUTHORS Starcich, B.R., Hahn, B.H., Shaw, G.M., McNeely, P.D., Modrow, S., Wolf, H., Parks, E.S., Parks, W.P., Josephs, S.F., Gallo, R.C. and Wong-Staal, F.

TITLE Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS

JOURNAL Cell 45 (5), 637-648 (1986)

MEDLINE 86218077

REFERENCE 21 (sites)

AUTHORS Willey, R.L., Rutledge, R.A., Dias, S., Folks, T., Theodore, T., Buckler, C.E. and Martin, M.A.

TITLE Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83 (14), 5038-5042 (1986)

MEDLINE 86259728

REFERENCE 22 (bases 8761 to 9060)

AUTHORS Fisher, A.G., Ratner, L., Mitsuya, H., Marselle, L.M., Harper, M.E., Broder, S., Gallo, R.C. and Wong-Staal, F.

TITLE Infectious mutants of HTLV-III with changes in the 3' region and markedly reduced cytopathic effects  
 JOURNAL Science 233 (4764), 655-659 (1986)  
 MEDLINE 86261824  
 REFERENCE 23 (sites)  
 AUTHORS Feinberg, M.B., Jarrett, R.F., Aldovini, A., Gallo, R.C. and Wong-Staal, F.

TITLE HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA  
 JOURNAL Cell 46 (6), 807-817 (1986)  
 MEDLINE 87002448  
 REFERENCE 24 (sites)  
 AUTHORS Lightfoote, M.M., Coligan, J.E., Folks, T.M., Fauci, A.S., Martin, M.A. and Venkatesan, S.

TITLE Structural characterization of reverse transcriptase and endonuclease polypeptides of the acquired immunodeficiency syndrome retrovirus  
 JOURNAL J. Virol. 60 (2), 771-775 (1986)  
 MEDLINE 87036947  
 REFERENCE 25 (sites)  
 AUTHORS Wright, C.M., Felber, B.K., Paskalis, H. and Pavlakis, G.N.

TITLE Expression and characterization of the trans-activator of HTLV-III/LAV virus  
 JOURNAL Science 234 (4779), 988-992 (1986)  
 MEDLINE 87042788  
 REFERENCE 26 (sites)  
 AUTHORS Terwilliger, E., Sodroski, J.G., Rosen, C.A. and Haseltine, W.A.

TITLE Effects of mutations within the 3' orf open reading frame region of human T-cell lymphotropic virus type III (HTLV-III/LAV) on replication and cytopathogenicity  
 JOURNAL J. Virol. 60 (2), 754-760 (1986)  
 MEDLINE 87036943  
 REFERENCE 27 (sites)  
 AUTHORS Goh, W.C., Sodroski, J.G., Rosen, C.A. and Haseltine, W.A.

TITLE Expression of the art gene protein of human T-lymphotropic virus type III (HTLV-III/LAV) in bacteria  
 JOURNAL J. Virol. 61 (2), 633-637 (1987)  
 MEDLINE 87112968  
 REFERENCE 28 (sites)  
 AUTHORS Modrow, S., Hahn, B.H., Shaw, G.M., Gallo, R.C., Wong-Staal, F. and Wolf, H.

TITLE Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions  
 JOURNAL J. Virol. 61 (2), 570-578 (1987)  
 MEDLINE 87112954  
 REFERENCE 29 (sites)  
 AUTHORS Muesing, M.A., Smith, D.H. and Capon, D.J.

TITLE Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein  
 JOURNAL Cell 48 (4), 691-701 (1987)  
 MEDLINE 87131081  
 REFERENCE 30 (sites)  
 AUTHORS Nabel, G. and Baltimore, D.

TITLE An inducible transcription factor activates expression of human immunodeficiency virus in T cells  
 JOURNAL Nature 326 (6114), 711-713 (1987)

MEDLINE 87173065  
 REMARK Erratum: [Nature 1990 Mar 8;344(6262):178]  
 REFERENCE 31 (sites)  
 AUTHORS Fisher,A.G., Ensoli,B., Ivanoff,L., Chamberlain,M., Petteway,S.,  
 Ratner,L., Gallo,R.C. and Wong-Staal,F.  
 TITLE The sor gene of HIV-1 is required for efficient virus transmission  
 in vitro  
 JOURNAL Science 237 (4817), 888-893 (1987)  
 MEDLINE 87292118  
 REFERENCE 32 (sites)  
 AUTHORS Patarca,R., Heath,C., Goldenberg,G.J., Rosen,C.A., Sodroski,J.G.,  
 Haseltine,W.A. and Hansen,U.M.  
 TITLE Transcription directed by the HIV long terminal repeat in vitro  
 JOURNAL AIDS Res. Hum. Retroviruses 3 (1), 41-55 (1987)  
 MEDLINE 87299195  
 REFERENCE 33 (sites)  
 AUTHORS Wong-Staal,F., Chanda,P.K. and Ghrayeb,J.  
 TITLE Human immunodeficiency virus: the eighth gene  
 JOURNAL AIDS Res. Hum. Retroviruses 3 (1), 33-39 (1987)  
 MEDLINE 87299194  
 REFERENCE 34 (bases 1 to 9635; 1 to 9635)  
 AUTHORS Ratner,L., Fisher,A., Jagodzinski,L.L., Mitsuya,H., Liou,R.S.,  
 Gallo,R.C. and Wong-Staal,F.  
 TITLE Complete nucleotide sequences of functional clones of the AIDS  
 virus  
 JOURNAL AIDS Res. Hum. Retroviruses 3 (1), 57-69 (1987)  
 MEDLINE 87299196  
 REFERENCE 35 (bases 6225 to 8795)  
 AUTHORS Reitz,M.S. Jr., Wilson,C., Naugle,C., Gallo,R.C. and  
 Robert-Guroff,M.  
 TITLE Generation of a neutralization-resistant variant of HIV-1 is due to  
 selection for a point mutation in the envelope gene  
 JOURNAL Cell 54 (1), 57-63 (1988)  
 MEDLINE 88253426  
 REFERENCE 36 (bases 790 to 2292)  
 AUTHORS Pal,R., Reitz,M.S. Jr., Tschachler,E., Gallo,R.C.,  
 Sarngadharan,M.G. and Veronese,F.D.  
 TITLE Myristoylation of gag proteins of HIV-1 plays an important role in  
 virus assembly  
 JOURNAL AIDS Res. Hum. Retroviruses 6 (6), 721-730 (1990)  
 MEDLINE 90303964  
 REFERENCE 37 (sites)  
 AUTHORS Ido,E., Han,H.P., Kezdy,F.J. and Tang,J.  
 TITLE Kinetic studies of human immunodeficiency virus type 1 protease and  
 its active-site hydrogen bond mutant A28S  
 JOURNAL J. Biol. Chem. 266 (36), 24359-24366 (1991)  
 MEDLINE 92105089  
 COMMENT On Mar 25, 1997 this sequence version replaced gi:327742.  
 [6] sites; tat mRNA and other transcript boundaries. [7] sites;  
 tat mRNA.  
 [8] sites; mRNA splice sites.  
 [9] sites; 27K antigen cds.  
 [5] sites; gp160 and gp120 coding sequences.  
 [1] sites; regulatory sequences in the LTR.  
 [(in) Weiss,R., Teich,N., Varmus,H. and Coffin,J. (Eds.); RNA Tumor  
 Viruses, Secon] review; bases 1 to 9718.  
 [15] sites; trans-activator function and TAR sequence. [19]

sites; pol coding sequence.  
 [22] sites; 23K sor gene product.  
 [23] sites; pol NH2-terminal region.  
 [20] sites; sor 23K protein.  
 [21] sites; sor 23K protein.  
 [24] sites; Spl binding sites in the promoter region. [17] sites; acceptor and donor splice sites for tat and 27K. [10] sites; deletion mutants in the tat gene.  
 [18] sites; env gene conserved/variable regions; separate entries.  
 [16] sites; trs cds boundaries.  
 [12] sites; trs cds boundaries.  
 [11] sites; env gene conserved/variable regions; separate entries.  
 [26] sites; tar or transactivator target.  
 [13] sites; 3' orf mutations.  
 [14] sites; pol p34 terminus.  
 [31] sites; promoter, TAR, tat-III mutants.  
 [32] sites; envelope protein epitopes.  
 [33] sites; trs/art protein.  
 [34] sites; inducible enhancer element.  
 [27] revises [30].  
 [29] sites; long terminal repeat.  
 [28] sites; R orf.  
 [35] sites; sor.

Sequence for [25] kindly provided in computer-readable form by L.Ratner, 19-AUG-1986.

The HXB2 sequence is being used as a reference genome for all the HIV entries because it has been derived from a demonstrably infectious clone. Hence not all of the 'sites' references above were concerned with this isolate.

FEATURES	Location/Qualifiers
source	1..9719 /organism="Human immunodeficiency virus type 1" /proviral /isolate="HXB2" /db_xref="taxon:11676" /note="HTLV-III/LAV"
LTR	1..634 /note="5' LTR"
repeat_region	454..551 /note="R repeat 5' copy"
mRNA	455..9635 /product="HXB2 genomic mRNA"
prim_transcript	455..9635 /note="tat, trs, 27K subgenomic mRNA"
intron	744..5777 /note="tat, trs, 27K mRNA intron 1"
CDS	790..2292 /note="gag polyprotein" /codon_start=1 /protein_id="AAB50258.1" /db_xref="GI:327745" /translation="MGARASVLSGGELDRWEKIRLRPGGKKKYKLKHIVWASRELERF AVNPGLLETSEGCRQILGQLQPSLQTGSEELRSLYNTVATLYCVHQRIEIKDKEALD KIEEEQNKSKKKAQQAADTGHNSQVSNYPYQNIQGMVHQAI SPRTLNAWVKVVE EKAFSPEVIMFMSALSEGATPQDLNLTMLNTVGGHQAAMQMLKETINEEAAEWDVHPV HAGPIAPGQMREPRGSDIAGTTSTLQEQIGWMTNNPPIPVGEIYKRWII LGLNKIVRM YSPTSILDIRQGPKEPFRDYVD R FYKTLRAEQASQEVKNWMTETLLVQNANPDCKTIL

KALGPAATLEEMMTACQGVGGPGHKARVLAEAMSQVTNSATIMMQRGNFRNQKIVKC  
 FNCGKEGHTARNCRAPRKKGCWKCGKEGHQMKDCTERQANFLGKIWPSYKGRPGNFLO  
 SRPEPTAPPEESFRSGVETTTTPQKQEPIDKELYPLTSLRSLFGNDPSSQ"  
 2358..5096  
 /note="pol polyprotein (NH2-terminus uncertain)"  
 /codon\_start=1  
 /protein\_id="AAB50259.1"  
 /db\_xref="GI:1906384"  
 /translation="MSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGP  
 TPVNIIGRNLLTQIGCTLNFPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEIC  
 TEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPH  
 PAGLKKKKSVTVLVDGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWK  
 GSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDLVVGSDLEIGQHRTKIEELRQHLLRW  
 GLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKLNWASQI  
 YPGIKVRQLCKLLRGTKALTEVIPLTEEALELAENREILKEPVHGVYDPSKDLIAE  
 IQKQGQGWTYQIYQEPFKNLKTGKYARMRGHAHTNDVKQLTEAVQKITTESIVIWGKT  
 PKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPLVLKWLQLEKEPIVGAETFYVDG  
 AANRETKLGKAGYVTNRGRQKVVTLTDTTNQKTELQAIYLALQDSGLEVNIVTDSQYA  
 LGIIQAQPDQSESELVNQIIIEQLIKKEKVYLAWVPAHKGIGGNEQVDKLVSAGIRKVL  
 FLDGIDKAQDEHEKYHSNWRAMASDFNLPPVVAKEIVASCDKQQLKGEAMHGQVDCSP  
 GIWQLDCTHLEGKVILVAVHVASGYIEAEVIPAETGQETAYFLLKLGRWPVKTIHTD  
 NGSNFTGATVRAACWWAGIKQEFGIYPNPQSQGVVESMNKELKKIIGQVRDQAEHLKT  
 AVQMAVFIHNFKRKGGIGGYSAGERIVDIIATDIQTKELQKQITKIQNFRVYRDSRN  
 PLWKGPAKLLWKGEAVVIQDNSDIKVVPRRKAKIIRDYGKQMGDDCVASRQDED"  
 5041..5619  
 /note="sor 23K protein"  
 /codon\_start=1  
 /protein\_id="AAB50260.1"  
 /db\_xref="GI:327747"  
 /translation="MENRWQVMIVWQVDRMRIRTWKSLVKHHMYVSGKARGWFYRHHY  
 ESPHPRISSEVHIPLGDARLVITTYWGLHTGERDWHLGQGVSIIEWRKRYSTQVDP  
 ADQLIHLYYFDCFSDSAIRKALLGHIVSPRCEYQAGHNKVGSLQYLALALALITPKKIK  
 PPLPSVTKLTEDRWKNPKQKTGHRGSHTMNGH"  
 5559..5795  
 /note="R (ORF) protein"  
 /codon\_start=1  
 /protein\_id="AAB50261.1"  
 /db\_xref="GI:327748"  
 /translation="MEQAPEDQGPQREPHNEWTLELLEELKNEAVRHFPRIWHLGLGQ  
 HIYETYGDTWAGVEAIRILQQLLFIHFQNWVST"  
 join(5831..6045,8379..8424)  
 /note="tat protein"  
 /codon\_start=1  
 /protein\_id="AAB50256.1"  
 /db\_xref="GI:1906383"  
 /translation="MEPVDPRLEPWKHPGSQPKTACTNCYCKKCCFHCQVCFITKALG  
 ISYGRKKRRQRRRAHQNSQTHQASLSKQPTSQPRGDPTGPKE"  
 5831..6045  
 /note="tat protein, first expressed exon"  
 /number=2  
 join(5970..6045,8379..8653)  
 /note="trs protein"  
 /codon\_start=1  
 /protein\_id="AAB50257.1"  
 /db\_xref="GI:327744"  
 /translation="MAGRSGDSDEELIRTVRLIKLLYQSNPPPNPEGTRQARRNRRRR  
 WRERQRQIHSISERILGTYLGRSAEPVPLQLPPLERLTLDNCNEDCGTSGTQGVGSPQI

```

exon          LVESPTVLES GTKE"
              5970..6045
              /note="trs protein, first expressed exon"
              /number=2
intron        6046..8378
              /note="tat, trs, 27K mRNA intron 2"
CDS           6225..8795
              /note="envelope polyprotein"
              /codon_start=1
              /protein_id="AAB50262.1"
              /db_xref="GI:1906385"
              /translation="MRVKEKYQHLWRWGWRWGTMLLGMLMICSATEKLWVTVYYGVPV
WKEATT'TLFCASDAKAYDTEVHNWVATHACVPTDPNPQEVVLNVNTENFNMWKNDMVE
QMHEDIISLWDQSLKPCVKLTPLCVSLKCTDLKNDTNTNSSSGRMIMEKGEIKNCSFN
ISTSIRGKVQKEYAFFYKLDIIPIDNDTTSYKLTSCNTSVITQACPKVSFEPIPIHYC
APAGFAILKCNKTFNGTGPCTINVSTVQCTHGIRPVVSTQLLLNGLSLAEFEVSVIRSVN
FTDNAKTIIVQLNTSVEINCTRPNNNTRKRIRIQRGPGRAFTIGKIGNMRQAHCNIS
RAKWNNTLTKQIASKLREQFGNNKTIIFKQSSGGDPEIVTHSFNCGGEFFYCNSTQLFN
STWFNSTWSTEGSNNTGSDTITLPCRICKIINMWQKVGKAMYAPPISGQIRCSSNIT
GLLLTRDGGNSNNESEIFRPGGGDMRDNRSELYKYKVVKIEPLGVAPTAKRRRVQR
EKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQNNLLRAIEAQQHLL
QLTVWGIKQLQARILAVEYLKDQQLLGIWGC SGKLICTTAVPWNASWSNKSLEQIWN
HTTWMEWDREINNYTSLIHSLEESQNQQEKNEQELLELDKWLWNWFNITNWLWYI
KLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGPDREGEIEEGGER
DRDRSIRLVNGSLALIWDLLRSLCLFSYHRLRDLILLIVTRIVELLGRRGWEALKYWWN
LLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQACRAIRHIPRRIRQGLERILL
"
exon          8379..8652
              /note="trs protein"
              /number=3
exon          8379..8424
              /note="tat protein"
              /number=3
CDS           8797..9168
              /note="27K protein (premature termination)"
              /codon_start=1
              /protein_id="AAB50263.1"
              /db_xref="GI:1906386"
              /translation="MGGKWSKSSVIGWPTVRERMRAEPAADRVGAA SRDLEKHGAIT
SSNTAATNAACAWLEAQEEEEVGFPVTPQVPLRPMTYKAAVDLSHFLKEKGGLEGLIH
SQRRQDILDWIYHTQGYFPD"
LTR           9086..9719
              /note="3' LTR"
repeat_region 9540..9636
              /note="R repeat 3' copy"
polyA_signal  9612..9617
              /note="HXB2 mRNA polyadenylation signal"

```

BASE COUNT      3411 a    1772 c    2373 g    2163 t

ORIGIN

```

1  tgggaagggt  aattcactcc  caacgaagac  aagatatacct  tgatctgtgg  atctaccaca
61  cacaaggcta  ctccctgat  tagcagaact  acacaccagg  gccagggatc  agatatccac
121 tgacctttgg  atggtgctac  aagctagtag  cagttgagcc  agagaagtta  gaagaagcca
181 acaaaggaga  gaacaccagc  ttgttacacc  ctgtgagcct  gcatggaatg  gatgaccocg
241 agagagaagt  gttagagtgg  aggtttgaca  gccgcctagc  atttcatac  atggcccag
301 agctgcatcc  ggagtacttc  aagaactgct  gacatcgagc  ttgctacaag  ggactttccg
361 ctggggactt  tccagggagg  cgtggcctgg  gcgggactgg  ggagtggcga  gccctcagat
421 cctgcatata  agcagctgct  ttttgccctg  actgggtctc  tctgggttag  ccagatctga

```



481	gcctgggagc	tctctggcta	actaggggaac	ccactgctta	agcctcaata	aagcttgctt
541	tgagtgtctt	aagtagtgtg	tgcccgctct	ttgtgtgact	ctggtaacta	gagatccctc
601	agaccctttt	agtcagtgtg	gaaaatctct	agcagtggcg	cccgaacagg	gacctgaaag
661	cgaaagggaa	accagaggag	ctctctcgac	gcaggactcg	gcttgctgaa	gcgcgcacgg
721	caagaggcga	ggggcggcga	ctggtgagta	cgccaaaaat	tttgactagc	ggaggctaga
781	aggagagaga	tgggtgcgag	agcgtcagta	ttaagcgggg	gagaattaga	tcgatgggaa
841	aaaattcggg	taaggccagg	gggaaagaaa	aaatataaat	taaaacatat	agtatgggca
901	agcaggggag	tagaacgatt	cgcagttaat	cctggcctgt	tagaaacatc	agaaggctgt
961	agacaaatac	tgggacagct	acaaccatcc	cttcagacag	gatcagaaga	acttagatca
1021	ttatataata	cagtagcaac	cctctattgt	gtgcatcaaa	ggatagagat	aaaagacacc
1081	aaggaagctt	tagacaagat	agaggaagag	caaaacaaaa	gtaagaaaaa	agcacagcaa
1141	gcagcagctg	acacaggaca	cagcaatcag	gtcagccaaa	attaccctat	agtgcagaac
1201	atccaggggc	aaatggtaca	tcaggccata	tcacctagaa	ctttaaatgc	atgggtaaaa
1261	gtagtagaag	agaaggcttt	cagcccagaa	gtgataccca	tgttttcagc	attatcagaa
1321	ggagccaccc	cacaagattt	aaacaccatg	ctaaacacag	tggggggaca	tcaagcagcc
1381	atgcaaatgt	taaaagagac	catcaatgag	gaagctgcag	aatgggatag	agtgcattcca
1441	gtgcatgcag	ggcctattgc	accaggccag	atgagagaac	caaggggaag	tgacatagca
1501	ggaactacta	gtacccttca	ggaacaaata	ggatggatga	caaataatcc	acctatccca
1561	gtaggagaaa	tttataaaa	atggataatc	ctgggattaa	ataaaatagt	aaagatgtat
1621	agccctacca	gcattctgga	cataagacaa	ggaccaagg	aacccttag	agactatgta
1681	gaccggttct	ataaaactct	aagagccgag	caagcttcac	aggaggtaaa	aaattggatg
1741	acagaaacct	tgttggtcca	aaatgcgaac	ccagattgta	agactatttt	aaaagcattg
1801	ggaccagcgg	ctacactaga	agaaatgatg	acagcatgtc	agggagtagg	aggacccggc
1861	cataaggcaa	gagttttggc	tgaagcaatg	agccaagtaa	caaattcagc	taccataatg
1921	atgcagagag	gcaattttag	gaaccaaaga	aagattgtta	agtgtttcaa	ttgtggcaaa
1981	gaagggcaca	cagccagaaa	ttgcaggggc	cctaggaaaa	agggctgttg	gaaatgtgga
2041	aaggaaggac	accaaataaa	agattgtact	gagagacagg	ctaatttttt	aggggaagatc
2101	tggccttcoo	acaagggaag	gccagggaat	tttcttcaga	gcagacagag	gccaacagcc
2161	ccaccagaag	agagcttcag	gtctggggta	gagacaacaa	ctccccctca	gaagcaggag
2221	ccgatagaca	aggaaactgta	tccttttaact	tccttcagg	caactcttgg	caacgacccc
2281	tcgtcacaat	aaagatagg	gggcaactaa	aggaagctct	attagataca	ggagcagatg
2341	atacagtatt	agaagaaatg	agtttgccag	gaagatggaa	acaaaaaatg	atagggggaa
2401	ttggagggtt	tatcaaagta	agacagtatg	atcagatact	catagaaatc	tgtggacata
2461	aagctatagg	tacagtatta	gtaggaccta	cacctgtcaa	cataattgga	agaaatctgt
2521	tgactcagat	tggttgcact	ttaaattttc	ccattagccc	tattgagact	gtaccagtaa
2581	aattaaagcc	aggaatggat	ggcccaaaa	ttaaacaatg	gccattgaca	gaagaaaaaa
2641	taaaagcatt	agtagaaatt	tgtacagaga	tggaaaagga	agggaaaatt	tcaaaaattg
2701	ggcctgaaaa	tccatacaat	actccagtat	ttgccataaa	gaaaaaagac	agtactaaat
2761	ggagaaaaat	agtagatttc	agagaactta	ataagagaac	tcaagacttc	tgggaagtcc
2821	aattaggaat	accacatccc	gcagggttaa	aaaagaaaaa	atcagtaaca	gtactggatg
2881	tgggtgatgc	atatttttca	gttcccttag	atgaagactt	caggaagtat	actgcattta
2941	ccatacctag	tataaacaat	gagacaccag	ggattagata	tcagtacaat	gtgcttccac
3001	agggatggaa	aggatcacca	gcaatatctc	aaagtagcat	gacaaaaatc	ttagagcctt
3061	ttagaaaaca	aaatccagac	atagttatct	atcaatacat	ggatgatttg	tatgtaggat
3121	ctgacttaga	aatagggcag	catagaacaa	aaatagagga	gctgagacaa	catctgttga
3181	ggtggggact	taccacacca	gacaaaaaac	atcagaaaga	acctccattc	ctttggatgg
3241	gttatgaaat	ccatcctgat	aaatggacag	tacagcctat	agtgtctgcca	gaaaaagaca
3301	gctggactgt	caatgacata	cagaagttag	tggggaaatt	gaattgggca	agtcagattt
3361	accaggggat	taaagtaagg	caattatgta	aactccttag	aggaacccaa	gcactaacag
3421	aagtaatacc	actaacagaa	gaagcagagc	tagaactggc	agaaaacaga	gagattctaa
3481	aagaaccagt	acatggagtg	tattatgacc	catcaaaaga	cttaatagca	gaaatacaga
3541	agcaggggca	aggccaatgg	acatatcaaa	tttatcaaga	gccattttaa	aatctgaaaa
3601	caggaaaata	tgcaagaatg	aggggtgccc	acactaatga	tgtaaaacaa	ttaacagagg
3661	cagtgcataa	aaataaccaca	gaaagcatag	taatatgggg	aaagactcct	aaatttaaac
3721	tgccatatac	aaaggaaaca	tgggaaacat	ggtggacaga	gtattggcaa	gccacctgga
3781	ttcctgagtg	ggagtttggt	aatacccttc	ccttagtgaa	attatgggtac	cagttagaga
3841	aagaacccat	agtaggagca	gaaaccttct	atgtagatgg	ggcagctaac	agggagacta

3901	aattagggaaa	agcaggatat	gttactaata	gaggaagaca	aaaagttgtc	accctaactg
3961	acacaacaaa	tcagaagact	gagttacaag	caatztatct	agctttgcag	gattcgggat
4021	tagaagtaaa	catagtaaca	gactcacaat	atgcattagg	aatcattcaa	gcacaaccag
4081	atcaaagtga	atcagagtta	gtcaatcaaa	taatagagca	gttaataaaa	aaggaaaagg
4141	tctatctggc	atgggtacca	gcacacaaa	gaattggagg	aatgaacaa	gtagataaat
4201	tagtcagtgc	tggaaatcagg	aaagtactat	ttttagatgg	aatagataag	gccaagatg
4261	aacatgagaa	atatcacagt	aattggagag	caatggctag	tgattttaac	ctgccacctg
4321	tagtagcaaa	agaaatagta	gccagctgtg	ataaatgtca	gctaaaagga	gaagccatgc
4381	atggacaagt	agactgtagt	ccaggaatat	ggcaactaga	ttgtacacat	ttagaaggaa
4441	aagttatcct	ggtagcagtt	catgtagcca	gtggatatat	agaagcagaa	gttattccag
4501	cagaaacagg	gcaggaaaca	gcatattttc	ttttaaaatt	agcaggaaga	tggccagtaa
4561	aaacaataca	tactgacaat	ggcagcaatt	tcaccgggtg	tacggttagg	gccgcctggt
4621	ggtgggcggg	aatcaagcag	gaatttggaa	ttccctacaa	tcccaaagt	caaggagtag
4681	tagaatctat	gaataaagaa	ttaaagaaaa	ttataggaca	ggtaagagat	caggctgaac
4741	atcttaagac	agcagtacaa	atggcagtat	tcattccaca	ttttaaaaga	aaagggggga
4801	ttggggggta	cagtgcaggg	gaaagaatag	tagacataat	agcaacagac	atacaacta
4861	aagaattaca	aaaacaaatt	acaaaaattc	aaaattttcg	ggtttattac	agggacagca
4921	gaaatccact	ttggaaagga	ccagcaaagc	tcctctggaa	aggtgaaggg	gcagtagtaa
4981	tacaagataa	tagtgacata	aaagtagtgc	caagaagaaa	agcaaagatc	attagggtatt
5041	atggaaaaca	gatggcagg	gatgattgtg	tggcaagtag	acaggatgag	gattagaaca
5101	tggaaaagtt	tagtaaaaca	ccatatgtat	gtttcaggga	aagctagggg	atggttttat
5161	agacatcact	atgaaagccc	tcattccaaga	ataagttcag	aagtacacat	cccactaggg
5221	gatgctagat	tggtataaac	aacatattgg	ggtctgcata	caggagaaag	agactggcat
5281	ttgggtcagg	gagtctccat	agaatggagg	aaaaagagat	atagcacaca	agtagaccct
5341	gaactagcag	accaactaat	tcattctgtat	tactttgact	gtttttcaga	ctctgtctata
5401	agaaaggcct	tattaggaca	catagttagc	cctagggtgtg	aatatcaagc	aggacataac
5461	aaggtaggat	ctctacaata	cttggcacta	gcagcattaa	taacacccaa	aaagataaag
5521	ccacctttgc	ctagtgttac	gaaactgaca	gaggatagat	ggaacaagcc	ccagaagacc
5581	aaggccacac	gaggagccca	cacaatgaat	ggacactaga	gcttttagag	gagcttaaga
5641	atgaagctgt	tagacatttt	cctaggattt	ggctccatgg	cttagggcaa	catatctatg
5701	aaacttatgg	ggatacttgg	gcaggagtgg	aagccataat	aagaattctg	caacaactgc
5761	tgtttatcca	ttttcagaat	tgggtgtcga	catagcagaa	taggcgttac	tcgacagagg
5821	agagcaagaa	atggagccag	tagatcctag	actagagccc	tggaaagcatc	cagggaagtca
5881	gcctaaaact	gcttgtacca	attgctattg	taaaaagtgt	tgctttcatt	gccaaagttg
5941	tttcataaca	aaagccttag	gcattctcta	tggcaggaag	aagcggagac	agcgacgaag
6001	agctcatcag	aacagtcaga	ctcatcaagc	ttctctatca	aagcagtaag	tagtacatgt
6061	aacgcaacct	ataccaatag	tagcaatagt	agcattagta	gtagcaataa	taatagcaat
6121	agttgtgtgg	tccatagtaa	tcataagaata	taggaaaata	ttaagacaaa	gaaaaataga
6181	cagggttaatt	gatagactaa	tagaaagagc	agaagacagt	ggcaatgaga	gtgaaggaga
6241	aatatcagca	cttgtggaga	tgggggtgga	gatggggcac	catgctcctt	gggatgttga
6301	tgatctgtag	tgctacagaa	aaattgtggg	tcacagtcta	ttatggggta	cctgtgtgga
6361	aggaagcaac	caccactcta	ttttgtgcat	cagatgctaa	agcatatgat	acagaggtac
6421	ataatgtttg	ggccacacat	gcctgtgtac	ccacagaccc	caaccacaaa	gaagtagtat
6481	tggtaaatgt	gacagaaaaa	tttaacatgt	ggaaaaatga	catggtagaa	cagatgcatg
6541	aggatataat	cagtttatgg	gatcaaagcc	taaagccatg	tgtaaaatta	acccactctt
6601	gtgttagttt	aaagtgcact	gatttgaaga	atgatactaa	taccaatagt	agtagcggga
6661	gaatgataat	ggagaaaagga	gagataaaaa	actgctcttt	caatatcagc	acaagcataa
6721	gaggtaaggt	gcagaaaagaa	tatgcatttt	tttataaaact	tgatataata	ccaatagata
6781	atgatactac	cagctataag	ttgacaagtt	gtaacacctc	agtcattaca	caggcctgtc
6841	caaaggtatc	ctttgagcca	attcccatac	attattgtgc	cccggctggg	tttgcgattc
6901	taaaatgtaa	taataagacg	ttcaatggaa	caggaccatg	tacaaatgtc	agcacagtac
6961	aatgtacaca	tggaaatagg	ccagtagtat	caactcaact	gctgttaaat	ggcagtctag
7021	cagaagaaga	ggtagtaatt	agatctgtca	atttcacgga	caatgctaaa	accataatag
7081	tacagctgaa	cacatctgta	gaaatlaatt	gtacaagacc	caacaacaa	acaagaaaaa
7141	gaatccgtat	ccagagagga	ccaggagag	catttgttac	aataggaaaa	ataggaaata
7201	tgagacaagc	acattgtaac	attagtagag	caaaatggaa	taacacttta	aaacagatag
7261	ctagcaaat	aagagaacaa	tttggaata	ataaaacaat	aatctttaag	caatcctcag

7321	gaggggaccc	agaaattgta	acgcacagtt	ttaattgtgg	aggggaattt	ttctactgta
7381	attcaacaca	actgtttaat	agtacttgg	ttaatagtag	ttggagtact	gaagggtcaa
7441	ataacactga	aggaagtga	acaatcacc	tcccatgcag	aataaaacaa	attataaaca
7501	tgtggcagaa	agtaggaaaa	gcaatgtatg	ccccctccat	cagtggacaa	attagatgtt
7561	catcaaatat	tacagggctg	ctattaacaa	gagatgggtg	taatagcaac	aatgagtcg
7621	agatcttcag	acctggagga	ggagatatga	gggacaattg	gagaagtga	ttatataaat
7681	ataaagtagt	aaaaattgaa	ccattaggag	tagcaccac	caaggcaag	agaagagtgg
7741	tgcaagagaga	aaaaagagca	gtgggaatag	gagctttgtt	ccttgggttc	ttgggagcag
7801	caggaagcac	tatgggcgca	gcctcaatga	cgctgacggt	acaggccaga	caattattgt
7861	ctggtatagt	gcagcagcag	aacaatttgc	tgagggtat	tgaggcgcaa	cagcatctgt
7921	tgcaactcac	agtctggggc	atcaagcagc	tccaggcaag	aatcctggct	gtggaaagat
7981	acctaaagga	tcaacagctc	ctggggattt	ggggttgctc	tggaaaaactc	atttgcacca
8041	ctgctgtgcc	ttggaatgct	agttggagta	ataaatctct	ggaacagatt	ttggaatcaca
8101	cgacctggat	ggagtgggac	agagaaatta	acaattacac	aagcttaata	cactccttaa
8161	ttgaagaatc	gcaaaaccag	caagaaaaga	atgaacaaga	attattggaa	ttagataaat
8221	gggcaagttt	gtggaattgg	tttaacataa	caaattggct	gtggtatata	aaattattca
8281	taatgatagt	aggaggcttg	gtaggtttaa	gaatagtttt	tgctgtactt	tctatagtga
8341	atagagttag	gcagggatat	tcaccattat	cgtttcagac	ccacctccca	accccgaggg
8401	gaccgcagag	gcccgaagga	atagaagaag	aaggtggaga	gagagacaga	gacagatcca
8461	ttcgattagt	gaacggatcc	ttggcactta	tctgggacga	tctgcggagc	ctgtgcctct
8521	tcagctacca	ccgcttgaga	gacttactct	tgattgtaac	gaggattgtg	gaacttctgg
8581	gacgcagggg	gtgggaagcc	ctcaaatatt	ggtggaatct	cctacagtat	tgagtcagg
8641	aactaaagaa	tagtgctgtt	agcttgctca	atgccacagc	catagcagta	gctgagggga
8701	cagatagggt	tatagaagta	gtacaaggag	cttgtagagc	tattcgccac	atacctagaa
8761	gaataagaca	gggcttgga	aggattttgc	tataagatgg	gtggcaagt	gtcaaaaagt
8821	agtgtgattg	gatggcctac	tgtaaggga	agaatgagac	gagctgagcc	agcagcagat
8881	aggggtgggag	cagcatctcg	agacctggaa	aaacatggag	caatcacaa	tagcaataca
8941	gcagctacca	atgctgcttg	tgcttgcta	gaagcacaag	aggaggagga	ggtgggtttt
9001	ccagtcacac	ctcaggtacc	tttaagacca	atgacttaca	aggcagctgt	agatcttagc
9061	cacttttttaa	aagaaaaggg	gggactggaa	gggctaattc	actcccaaag	aagacaagat
9121	atccttgatc	tgtggatcta	ccacacacaa	ggctacttcc	ctgattagca	gaactacaca
9181	ccagggccag	gggtcagata	tccactgacc	tttggatgg	gctacaagct	agtaccagtt
9241	gagccagata	agatagaaga	ggccaataaa	ggagagaaca	ccagcttgtt	acaccctgtg
9301	agcctgcatg	ggatggatga	ccggagagaa	gaagtgttag	agtggagggt	tgacagccgc
9361	ctagcatttc	atcacgtggc	ccgagagctg	catccggagt	acttcaagaa	ctgctgacat
9421	cgagcttgct	acaagggact	ttccgctggg	gactttccag	ggaggcggtg	cctgggcggg
9481	actggggagt	ggcgagccct	cagatcctgc	atataagcag	ctgctttttg	cctgtactgg
9541	gtctctctgg	ttagaccaga	tctgagcctg	ggagctctct	ggctaactag	ggaaccact
9601	gcttaagcct	caataaagct	tgcttgagt	gcttcaagta	gtgtgtgccc	gtctgtgtg
9661	tgactctggt	aactagagat	ccctcagacc	cttttagtca	gtgtggaaaa	tctctagca

The oligonucleotide probe for Ap-1 sequence in step 63 was synthesized using PMA responsive element as consensus sequence as indicated by the reference of Northrop et al., 1993, and adding flanking sequences.

These two sequences which were used as probes are representative example to demonstrate the methodology of DNA-protein interaction. Any other relevant sequence(s) can be used for this purpose.